The Diffusion of Viral Content in Multi-layered Social Networks

Jarosław Jankowski[1], Michał Kozielski[2], Wojciech Filipowski[2], and Radosław Michalski[3]

[1] Faculty of Computer Science, West Pomeranian University of Technology, Szczecin, Poland
jjankowski@wi.zut.edu.pl
[2] Faculty of Automatic Control, Electronics & Computer Science, Silesian University of Technology, Gliwice, Poland
{michal.kozielski,wojciech.filipowski}@polsl.pl
[3] Institute of Informatics, Wrocław University of Technology, Poland
radoslaw.michalski@pwr.wroc.pl

Abstract. Modelling the diffusion of information is one of the key areas related to activity within social networks. In this field, there is recent research associated with the use of community detection algorithms and the analysis of how the structure of communities is affecting the spread of information. The purpose of this article is to examine the mechanisms of diffusion of viral content with particular emphasis on cross community diffusion.

Keywords: diffusion of information, multi-layered social networks, clustering algorithms, multiplex networks, social network analysis.

1 Introduction

Social platforms have become one of the leading trends in the development of the World Wide Web. Due to their specifics and worldwide reach with huge number of users, they create new challenges in the area of analytical methods for processing large data sets. They are used for social analysis and decision support processes as well as marketing activities. The methods in the area of social network analysis are focused on different fields ranging from static analysis towards network dynamics [15]. One of the growing areas is the analysis of information diffusion with focus on diffusion models [11], analysis of the factors affecting the dynamics of the spread of information [13] and optimization of these processes as well as maximization of the reach [9]. These methods are used in marketing analysis [1] as well as for monitoring of users' activity and the flow of digital content [3]. In recent years more attention is paid to the usage of community detection algorithms oriented to the diffusion of information [2] and methods oriented to maximization of the information flow within and between communities [4].

This article attempts to clarify the mechanisms of diffusion of virtual goods in the multiplayer platform functioning as a virtual world in which users can exchange graphics, avatars or different content in viral messages. Users of such an environment

C. Bădică, N.T. Nguyen, and M. Brezovan (eds.): ICCCI 2013, LNAI 8083, pp. 30–39, 2013.

create a multi-layered network, where each layer can be defined with different user activities. Analysis of such a network is an interesting but challenging problem [6]. In general, each layer of multi-layer social network can consist of different set of actors connected by different relations. Therefore, the following issues have to be considered when such complex structure is analysed: the communities identified in different layers do not have to fit to each other, some of them can be present in several layers, others can be formed only in one layer. Therefore, there are several issues concerning, e.g., common or separate layers analysis, layers compatibility or significance. These issues influencing the community identification have also impact on the analysis of information diffusion when multi-layered network structure is considered.

The goal of this work is to examine the mechanisms of diffusion of virtual goods with particular emphasis on cross community diffusion, where the communities were identified in multi-layered social network. Another goal is to verify whether the communities identified in different network layers can affect the information diffusion process. Finally, the work aims to give additional insight into the domain of information diffusion in social networks by the analysis of three real life datasets.

2 Related Work

The spread of information has become one of the main areas of the study of the phenomena associated with social media platforms. Research has been carried out on the flow of information in the systems, modelling social influence [14], studying the dynamics of viral campaigns and searching methods of optimization. Research on the theoretical and generalized mechanisms of diffusion processes including its application areas as well as the search for factors affecting the range and other parameters to evaluate their effectiveness [1] is performed.

Recently, more and more analysis related to the characteristics of the diffusion mechanisms is based on algorithms from the field of social network analysis. The static network parameters [1], the variability of the network at the time [15] and the attributes of the participants in the processes of diffusion and structures of interactions [10] are taken into account. The study of diffusion mechanisms on the basis of available influence models is focused on maximizing the range [11], selection of seeding nodes (e.g. [12]) and the analysis of the factors and parameters of social networks which affect the range [1]. The study also draws attention to the dependence of the flow of information related to the structure of the communities identified within the network. There are community detection algorithms in use which are associated with the identification of structures based on various definitions of groups [8], which can be used in networks analysis and for single as well as multi-layer networks [6]. What is recently undertaken by researchers is to try to connect the two areas and communities overlapping detection based on data on the diffusion with the assumption that the information propagation and relationships within communities are closely related [2]. Earlier experiences showed that clusters limit cascades of information and analysis of cluster structures can be a base for adjusting strategies to network structures and detected communities [7]. Other related research emphasized the ability to use interactions and the content of the transmitted messages in finding overlapping communities

[10]. Detecting communities can be used for influence maximization problems and selection of influential nodes [14]. Apart from community detection, research was performed on the flow of information inside and between communities. Integration of community detection algorithms by modelling the spread of information makes it possible to detect influence not only between individuals but between groups which can be useful for large systems where more importantly it is the global view than analysis at the micro scale [3]. Other attempts in this field are oriented towards maximizing diffusion by targeting online communities and increasing cross community communication [4].

While social platforms gather much more information than graph structures, some of the possible directions is using more information sources. An extended analysis based on communication topics, graphs topology and level of participation in communities measured by a number of interactions was performed recently [10]. The proposed method of community discovery using fuzzy modularity integrates several sources of information.

The literature review shows that cross community diffusion is one of the latest topics in this research field and still has open research challenges. The diffusion of information in the social networks is analysed in the context of communities and is usually based on information flow. In this work authors targeted the research to viral content diffusion in online community functioning in a form of virtual world in multi-layered network with the aim to determine what is the mechanism of the in different communities identified within different layers.

3 Theoretical Background and Dataset Description

The idea of multi-layered social network, sometimes called a multiplex network, assumes that there exists more than one communication type between users, e.g. between two particular nodes in the directed network there may be formed at least two edges of different type from one node to another [6]. By separating these types of links into different layers, a multi-layered social network is built. An example of multi-layered social network may be telecom network, where one layer is the layer of phone calls, the second one represents text messages and the third one – multimedia messages. What is natural, these layers may consist of different nodes and edges. This sort of networks allows to analyse layers separately or combined, depending on the desired goal. An example of multi-layered network is presented in Figure 1.

Another concept used in the latter part of this paper are clusters in social networks. Typically, a cluster is considered as a structure, in which network nodes are joined together in tightly knit groups, between which there are only looser connections [8].

The Louvain method [5] was applied to community identification in the work presented. It is based on the concept of modularity maximization and consists of two phases. The algorithm is reminiscent of the self-similar nature of complex networks and naturally incorporates a notion of hierarchy, as communities of communities are built during the process [5]. The algorithm has the computational complexity of O(m), where m is the number of nodes. The output of the method are non-overlapping groups, what enables the analysis of intra- and inter-network communication.

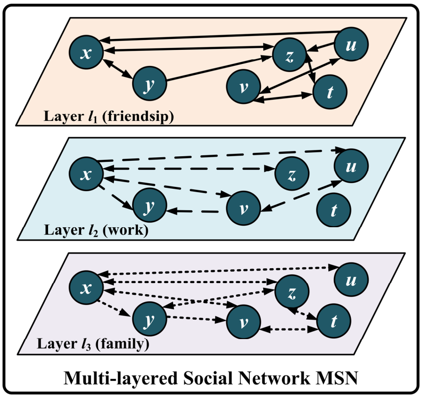


Fig. 1. An example of multi-layered social network [6]

Previous studies indicate a relationship between the processes of diffusion of information and the structures of identified communities within social networks. This area draws attention in recent publications and not all mechanisms are explained yet. In the present study, the aim was to analyse the mechanisms of diffusion in groups for different actions based on viral marketing.

This research is based on the real-life datasets presented in the following paragraphs. The analysis required the extraction of data from viral campaigns carried out in a single environment with comparable results with the ability of monitoring the media diffusion. For the purpose of verification mechanisms in the study, data from the viral campaigns (actions) denoted as V_1, V_2 and V_3 were used. The campaigns were conducted within the virtual world system in which users communicate in a synchronous chat system and their representation in the form of visual avatars. These three campaigns are described later in this section. Within the system, users can transmit messages to the other users and tracing of diffusion processes is possible.

In addition to the social network structure, a number of messages was used to estimate the relationship between a sender and a receiver. Weights W_{AB} and W_{BA} between user A and user B was represented by the number of private messages sent between them from the beginning of the relationship in both directions.

The first viral action denoted as V_1 was completed in one day. Five randomly selected users within seeding received new avatars that can be sent to other users. The mechanism of transmission of viral content did not require connection within friends list between sender and receiver.

The second action V_2 was associated with the users organizing a protest against the sharpening of legal regulations in the field of electronic media. The campaign members have the ability to upload graphical components associated with the action. The transmission mechanism of social relationships required communication between a sender and a receiver in the form of relation on the list of friends. A detailed monitoring of the media diffusion was possible and the action lasted for five days.

Action V_3 was associated with competitions in which winners received gifts and were able to give them to friends. Mechanics of diffusion required the presence of the receiver on the sender's friends list. The values characterising each campaign are presented in Table 1.

Table 1. Viral campaign characteristics

	Unique senders (A)	Unique receivers (B)	Infections sent per A	W_{AB}	W_{BA}
V_1	82	474	5.78	4.15	5.45
V_2	86	224	2.6	13.93	13.35
V_3	218	627	2.88	12.91	15.34

Preliminary analysis of the data indicates the presence of a stronger relationship between the sender and the recipient in action V_2 and V_3 than in the V_1 in action. The campaign V_1 achieved a higher rate of infection transmitted by a single user, however, it was relatively low conversion rate from receiver to sender at 0.17 while for V_2 and V_3 this factor was at 0.38 and 0.35 respectively. In all the analysed actions, senders had a higher login rate than receivers. For the action V_1 login ration of sender in relation to receiver was 1.59, in actions V_2 and V_3 was 1.8 and 1.9 respectively. Such differences indicate potential exposure of the sender as a user with a particular reputation for customer with less experience in the system usage. As shown in the preliminary analysis conducted, the actions have different characteristics, but the most noticeable differences are between the action V_1 and actions V_2 and V_3. In this context, analyses of the diffusion in communities and determining how the message was distributed within the communities can bring extensions to current research in this field.

The social network characterised above can be presented as consisting of two layers. Social layer is based on the relations derived from a list of friends of each user. Communication layer is based on the relations derived from a communication activity of each user. A message sent by a user during a viral action is represented in this case as a relation between the users. The experiments presented in the next paragraph enable the analysis the diffusion mechanisms in groups identified in these two layers.

4 Experimental Results

The main part of the study included the identification and analysis of the mechanisms of spread of information within and between communities. The results obtained show the different characteristics of the various communities within the social (Table 2) and communication (Table 3) network. The tables 2 and 3 present characteristics of the main communities identified. These characteristics cover a number of infections between the groups $C_{i,j}$, percentage of within group infections and outgoing infections and the respective percentage values.

In the social network (Table 2) for V_1 the largest community $C_{1,1}$ consists of 1,640 members and a lowest number of members at level of 277 was in community $C_{1,11}$. Most logins per user occurred in the community $C_{1,\ 19}$ which indicates the most advanced users in this group. Community $C_{1,5}$ gathered users with low activity. The largest community for V_2 was identified as $C_{2,2}$ and it consists of 1,477 members. The highest average number of logins was in group $C_{2,0}$ at the level of 835 logins per user. The three main communities associated with V_3 have the highest number of members. In community $C_{3,4}$ 3486 users were identified. The highest number of logins at 583 was detected in community $C_{3,5}$.

Analysis of the communication network (Table 3) shows the highest number of 260 users for V_1 campaign in the community $C_{1,15}$. For campaign V2 and V3 communities $C_{2,2}$ and $C_{3,3}$ got 198 and 945 users respectively. Tables 2 and 3 show the number of internal and external transmission between groups in two networks – social and communication ones.

Table 2. Intergroup and intragroup infections – Social Network

V_i		$C_{1,2}$	$C_{1,9}$	$C_{1,11}$	$C_{1,19}$	G_{IN}	G_{OUT}	SUM_{IN}	SUM_{OUT}
V_1	$C_{1,2}$	28	9	0	17	51,85%	48,15%	28	26
	$C_{1,9}$	16	28	2	27	21,92%	78,08%	28	45
	$C_{1,11}$	0	1	0	1	0,00%	100,00%	0	2
	$C_{1,19}$	101	60	13	85	39,00%	61,00%	85	174
	Total:	145	98	15	130	36,34%	63,66%	141	247
		$C_{2,0}$	$C_{2,1}$	$C_{2,2}$	$C_{2,3}$	G_{IN}	G_{OUT}	SUM_{IN}	SUM_{OUT}
V_2	$C_{2,0}$	31	5	23	15	41,89%	58,11%	31	43
	$C_{2,1}$	1	1	3	0	20,00%	80,00%	1	4
	$C_{2,2}$	25	6	39	13	30,12%	69,88%	39	44
	$C_{2,3}$	6	1	14	12	18,18%	81,82%	12	21
	Total:	63	13	79	40	42,56%	57,44%	83	112
		$C_{3,1}$	$C_{3,4}$	$C_{3,5}$		G_{IN}	G_{OUT}	SUM_{IN}	SUM_{OUT}
V_3	$C_{3,1}$	32	46	15		34,41%	65,59%	32	61
	$C_{3,4}$	56	139	58		22,13%	77,87%	139	114
	$C_{3,5}$	17	43	59		14,29%	85,71%	59	60
	Total:	105	228	132		49,46%	50,54%	230	235

The results for the infection between the groups indicated that not all campaigns behaved similarly. The campaign V_1 had more infections between groups, probably because of the impact of the mechanics of the campaign which did not require the existence of receivers on the friends list. While the number of external infections was high, at the same time infections did not reach the group $C_{1,5}$ which could be due to the isolation of the group and a major share of the beginners with low engagement in this type of actions. In campaigns V_2 and V_3 more inter-group infections were observed. A large number of infections between the groups in the observed campaigns was clearly associated with a high density network. Fig. 2 shows infections within the groups and between the groups and their variability over time. The x axis indicated the next steps determined by the individual campaigns infections. Dotted line shows an increase in external infections and continuous - internal infections within groups.

Table 3. Intergroup and intragroup infections – Communication Network

V_i		$C_{1,3}$	$C_{1,5}$	$C_{1,6}$	$C_{1,15}$	G_{IN}	G_{OUT}	SUM_{IN}	SUM_{OUT}
V_1	$C_{1,3}$	1	0	0	2	33.33%	66.67%	1	2
	$C_{1,5}$	0	3	6	2	27.27%	72.73%	3	8
	$C_{1,6}$	1	1	3	3	37.50%	62.50%	3	5
	$C_{1,15}$	25	25	41	111	54.95%	45.05%	111	91
	Total:	27	29	50	118	52.68%	47.32%	118	106
		$C_{2,2}$	$C_{2,3}$	$C_{2,4}$	$C_{2,5}$	G_{IN}	G_{OUT}	SUM_{IN}	SUM_{OUT}
V_2	$C_{2,2}$	5	0	4	0	55.56%	44.44%	5	4
	$C_{2,3}$	12	80	6	9	74.77%	25.23%	80	27
	$C_{2,4}$	1	3	16	1	76.19%	23.81%	16	5
	$C_{2,5}$	3	4	1	12	60.00%	40.00%	12	8
	Total:	21	87	27	22	71.97%	28.03%	113	44
		$C_{3,2}$	$c_{3,3}$	$c_{3,5}$	$C_{3,12}$	G_{IN}	G_{OUT}	SUM_{IN}	SUM_{OUT}
V_3	$C_{3,2}$	9	0	3	3	60.00%	40.00%	9	6
	$C_{3,3}$	2	19	8	2	61.29%	38.71%	19	12
	$C_{3,5}$	16	12	161	34	72.20%	27.80%	161	62
	$C_{3,12}$	15	21	44	205	71.93%	28.07%	205	80
	Total:	42	52	216	244	71.12%	28.88%	394	160

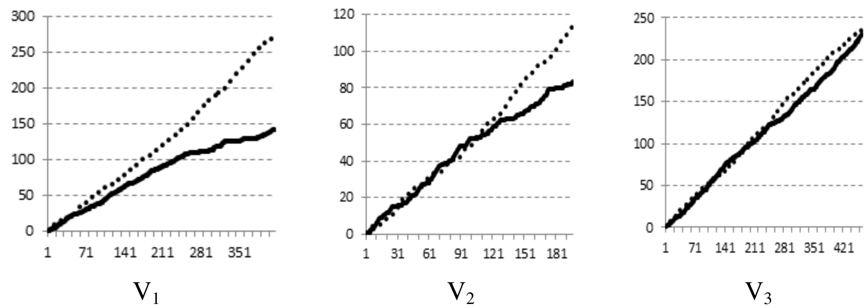
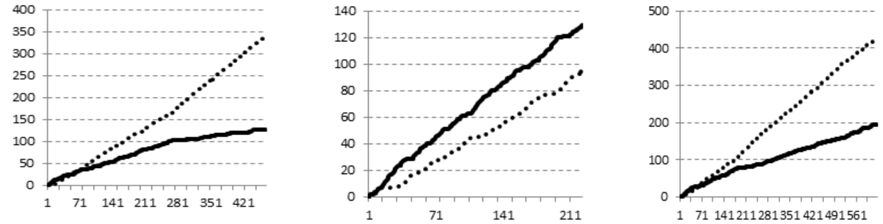


Fig. 2. Inter- and intra-group infections (dotted and solid line respectively) for campaigns V_1, V_2 and V_3 for social communities

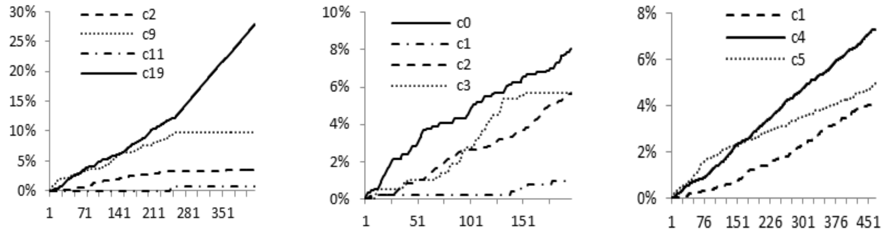
The results indicate that the infections between groups had the largest share in the campaign V_1. The V_3 campaign was characterized by a highest proportion of intergroup infections. Other relation between inter and intra group infections can be observed for groups detected in the communication layer what is presented in the Fig. 3.



V_1 V_2 V_3

Fig. 3. Inter- and intra-group infections (dotted and solid line respectively) for campaigns V1, V2 and V3 for communication communities

The next step examines growth in the number of infections in the individual groups and communities based on saturation measures. In Fig. 4 the growth rate of infection in each group and saturation is showed.



V_1 V_2 V_3

Fig. 4. Saturation in communities in social layer for viral content in V_1, V_2 and V_3 campaigns

The results indicate that for certain groups a relatively quick stabilization of infections was observed. For example, for V_1 the greatest saturation was achieved for community $C_{1,19}$ of up to 30% and dynamic of infections was highest in that community. The community $C_{1,9}$ despite the initial growth comparable to $C_{1,19}$ has stabilized at 10%. Dynamics of saturation in the campaign V_2 indicate the possibility of achieving a critical mass and a sharp increase at some point, which could be observed for the community $C_{2.3}$. Changes of dynamics in infections were observed within campaign V_3, where $C_{3,5}$ initially ran in the highest rates but after infection, 151 higher infection rates was obtained for community $C_{3,4}$. In Fig. 5 the growth rate of infection in each group for communication layer and saturation is showed.

For the communication network layer in the V_1 action the highest level of saturation was obtained for community $C_{1,3}$ and it was at two times higher level than in the social network layer. Similar pattern and relation was observed among well infected communities and those with lower saturation. For the second campaign in the communication network a single main community was identified while in the social network the communities were saturated at the similar levels. Within third campaign saturation in both network layers had similar patterns for two main communities.

Results showed differences between network layers and higher saturation in temporal communication layer.

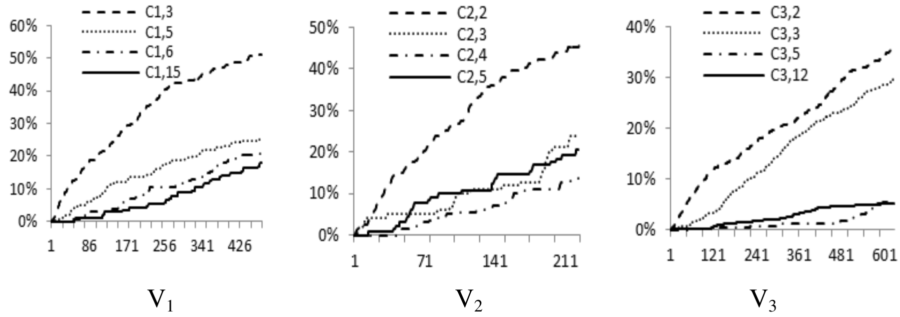


Fig. 5. Saturation in communities in communication layer for viral content in V_1, V_2 and V_3 campaigns

5 Conclusions and Future Work

The study focused on the existence of different mechanisms of diffusion based on two factors - network parameters and characteristics of users. In the presence of a weak relationship between the sender and the recipient, diffusion occurs in a wider range and increased intergroup infections can be observed. However, the spread between groups does not necessarily determine the success of the action (see V_1 action and $C_{1,5}$ community).

The results showed that communities identified in different layers did not fit to each other and different diffusion mechanisms were observed in each later on. Social network treated as more stable over the time showed higher intra group infections than within communities identified in the communication network. A relatively high density of analysed networks showed that in such a case intra community diffusion can be observed at acceptable level. Only at a very early stage, infections were propagated within the community. Next, inter community infections were observed. The performed analysis did not show compatibility of communities between layers and the community overlapping rate at both layers is low. However, for campaign V_1 the same proportion of inter and intra community infections was observed in both layers.

The results show that during viral actions targeted to communities deciding on which layer to use is more important. In the presented examples, if the campaign is addressed in a short term results would be better to target communities in the communication layer while for longer term results with higher impact on the whole community targeting communities in social layer would be more adequate. The analysis of cross community diffusion showed that temporal communication network and its specification was a better environment for inter group infections than stable network.

As a part of future works, the diffusion prediction in different network segments and the selection of seeding strategies oriented to communities can be considered. Additionally, other community identification methods are planned to be verified.

Acknowledgments. The research was partially supported by the European Commission under the 7th Framework Programme, Coordination and Support Action, Grant Agreement Number 316097, ENGINE - European research centre of Network intelliGence for INnovation Enhancement (http://engine.pwr.wroc.pl/) and the fellowship co-financed by European Union within European Social Fund.

References

1. Bampo, M., et al.: The Effects of the Social Structure of Digital Networks on Viral Marketing Performance. Information Systems Research 19(3), 273–290 (2008)
2. Barbieri, N., Bonchi, F., Manco, G.: Cascade-based Community Detection. In: Proceedings of the 6th ACM International Conference on Web Search and Data Mining (WSDM 2013). ACM, Rome (2013)
3. Belák, V., Lam, S., Hayes, C.: Cross-Community Influence in Discussion Fora. In: ICWSM (2012)
4. Belák, V., Lam, S., Hayes, C.: Towards Maximising Cross-Community Information Diffusion. In: Proceedings of ASONAM 2012, pp. 171–178 (2012)
5. Blondel, V.D., et al.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 10, P10008 (2008)
6. Bródka, P., Kazienko, P., Musiał, K., Skibicki, K.: Analysis of Neighbourhoods in Multi-layered Dynamic Social Networks. International Journal of Computational Intelligence Systems 5(3), 582–596 (2012)
7. Easley, D., Kleinberg, J.: Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press (2010)
8. Fortunato, S.: Community detection in graphs. Phys. Rep. 486(3-5), 75–174 (2010)
9. Goyal, A., et al.: On Minimizing Budget and Time in Influence Propagation over Social Networks. In: Social Network Analysis and Mining. Springer (2012)
10. Jankowski, J., Michalski, R., Kazienko, P.: The Multidimensional Study of Viral Campaigns as Branching Processes. In: Aberer, K., Flache, A., Jager, W., Liu, L., Tang, J., Guéret, C. (eds.) SocInfo 2012. LNCS, vol. 7710, pp. 462–474. Springer, Heidelberg (2012)
11. Kempe, D., Kleinberg, J.M., Tardos, É.: Maximizing the spread of influence through a social network. In: KDD, pp. 137–146 (2003)
12. Ma, H., Yang, H., Lyu, M.R., King, I.: Mining social networks using heat diffusion processes for marketing candidates selection. In: Proceedings of the 17th ACM Conf. on Information and Knowledge Management, pp. 233–242 (2008)
13. Najar, A., Denoyer, L., Gallinari, P.: Predicting information diffusion on social networks with partial knowledge. In: WWW, pp. 1197–1204 (2012)
14. Wang, Y., et al.: Community-based greedy algorithm for mining top-K influential nodes in mobile social networks. In: ACM SIGKDD 2010, pp. 1039–1048. ACM, New York (2010)
15. Watts, D., Strogatz, S.: Collective dynamics of ‘small-world’ networks. Nature 393, 440–442 (1998)